

ABSTRACT

A DTD and pattern tree creating module expresses in a tree form a hierarchical structure of respective elements defined by DTD and pattern information, and adds a description pattern specified with respect to the element concerned to each node in the tree. An entire control module requests a pattern retrieving module to retrieve based on the specified description pattern for every node in this tree. The pattern retrieving module extracts a region coincident with the specified description pattern out of a processing target document, and sends the region back to the entire control module. The entire control module adds tags corresponding to the element in front and rear of the region of a text that has been sent back as what corresponds to the element, thereby outputting a structured document.